

The P600-as-P3 hypothesis revisited: Single-trial analyses reveal that the late EEG positivity following linguistically deviant material is reaction time aligned



Jona Sassenhagen^{a,b}, Matthias Schlesewsky^b, Ina Bornkessel-Schlesewsky^{a,c,*}

^a Department of Germanic Linguistics, University of Marburg, Marburg, Germany

^b Department of English and Linguistics, Johannes Gutenberg-University, Mainz, Germany

^c School of Psychology, Social Work and Social Policy, University of South Australia, Adelaide, Australia

ARTICLE INFO

Article history:

Accepted 27 July 2014

Available online 24 August 2014

Keywords:

P600

P3

Single-trial analysis

Sentence processing

Syntax

Semantics

N400

Attention

Reorienting

Locus Coeruleus

ABSTRACT

The P600, a late positive ERP component following linguistically deviant stimuli, is commonly seen as indexing structural, high-level processes, e.g. of linguistic (re)analysis. It has also been identified with the P3 (P600-as-P3 hypothesis), which is thought to reflect a systemic neuromodulator release facilitating behavioural shifts and is usually response time aligned. We investigated single-trial alignment of the P600 to response, a critical prediction of the P600-as-P3 hypothesis. Participants heard sentences containing morphosyntactic and semantic violations and responded via a button press. The elicited P600 was perfectly response aligned, while an N400 following semantic deviations was stimulus aligned. This is, to our knowledge, the first single-trial analysis of language processing data using within-sentence behavioural responses as temporal covariates. Results support the P600-as-P3 perspective and thus constitute a step towards a neurophysiological grounding of language-related ERPs.

© 2014 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-SA license (<http://creativecommons.org/licenses/by-nc-sa/3.0/>).

1. Introduction

Linguistic deviations, especially of a structural or “rule-based” nature, often elicit a P600: averaged EEG recordings show a broad vertex-positive deflection beginning 500–1000+ ms after deviant word onset, typically with a centro-parietal maximum. In spite of extensive research into its antecedents, considerable disagreement remains about the neurophysiology underlying the P600.

Upon its discovery (Osterhout & Holcomb, 1992; see also Hagoort, Brown, & Groothusen, 1993), the P600 was seen as a new, distinct component reflecting aspects of combinatorial processing, e.g. the resolution of syntactic errors. Today, many researchers consider the P600 a specific component reflecting interpretative/integrative brain processes (e.g. Brouwer, Fitz, & Hoeks, 2012; Friederici, 2011; Gouvea, Phillips, Kazanina, & Poeppel, 2010; Kaan, 2007; Osterhout & Hagoort, 1999). Others (e.g. Bornkessel-Schlesewsky et al., 2011; Coulson, King, & Kutas,

1998a; Münte, Heinze, Matzke, Wieringa, & Johannes, 1998; van de Meerendonk, Kolk, Vissers, & Chwilla, 2010; Vissers, Kolk, van de Meerendonk, & Chwilla, 2008) view the P600 as a P3b, an instance of the well-known P3 component family.

Here, we approach the P600/P3 discussion from a novel perspective. By applying single-trial ERP analyses to a P600-eliciting paradigm, we aimed to test whether the P600 shows a well-established property of the P3: latency alignment with reaction times. We argue that, if the response properties between the P600 and P3 are similar in this respect, this strengthens the view that we can draw upon the wealth of existing knowledge about the psychological and neural properties of the P3 to inform a detailed, neurobiologically grounded view of the P600.

1.1. The P3¹ and response alignment

Like the P600, the P3 is a broad positive wave, often with a centro-parietal maximum. It is elicited anywhere from 250 to 1000+ ms after motivationally significant events. The best-known

* Corresponding author at: School of Psychology, Social Work and Social Policy, University of South Australia, GPO Box 2471, Adelaide, SA 5001, Australia. Fax: +61 (0)8 8302 4377.

E-mail address: Ina.Bornkessel-Schlesewsky@unisa.edu.au (I. Bornkessel-Schlesewsky).

¹ Unless otherwise noted, we use ‘P3’ to refer to the ‘P3b’ subcomponent of the P3 family throughout this paper.

paradigm for eliciting P3 effects is the oddball paradigm, in which participants engage in a task involving infrequent target stimuli amongst frequent standard stimuli (i.e. targets are responded to, counted etc.). Accordingly, the P3 is often described as a component that is elicited by uncertain, unexpected or surprising stimuli (e.g. Donchin, 1981; Sutton, Braren, Zubin, & John, 1965). However, while unexpectedness constitutes a very effective way of rendering a stimulus subjectively significant, it is neither a sufficient nor a necessary precondition. For example, task-relevant stimuli (i.e. stimuli which require a response) engender a higher P3 amplitude than stimuli which do not, even when stimulus frequency is equated between the two stimulus categories (Duncan-Johnson & Donchin, 1977). A P3 also follows significant or intrusive stimuli in fully task-free paradigms, e.g. to one's own name even while asleep or comatose (Perrin et al., 1999, 2006), and non-task relevant stimuli of personal significance during standard psychological tasks, like one's own cellphone ringtone (Roye, Jacobsen, & Schröger, 2007) or name (Gray, Ambady, Lowenthal, & Deldin, 2004) as a distractor item. Thus, the P3 does not reflect absolute probability or inherent salience, but subjective salience, and it is reduced or absent for unattended stimuli (Spencer, Dien, & Donchin, 2001).

A wide range of proposals has been put forward in order to account for the cognitive and functional significance of the P3. These include the influential Context Updating account (Donchin, 1981; Donchin & Coles, 1988; see also Polich, 1985, 2007), according to which the P3 reflects memory adaptations following critical events. Another prominent account (Verleger, 1988; Verleger, Jaśkowski, & Wascher, 2005) assigns a more tactical role to the P3 by proposing that it marks the linkage between critical events and reactions (henceforth: the Linking account of the P3). In a more strongly biologically-grounded approach, Nieuwenhuis, Aston-Jones and Cohen (2005) associate the P3 with the norepinephrine (NE) neuromodulator system and systemic NE release from the brainstem nucleus Locus Coeruleus (LC), which facilitates general cortical state transitions and thus supports cognitive reorientation (like response execution or inhibition). All approaches agree that the P3 follows highly salient events such as novel and unexpected events, highly task-relevant expected events, and self-relevant stimuli. In contrast to the Context Updating theory, however, the Linking and LC/NE accounts stress that, if a task requires overt behaviour and elicits a P3, there is a tight temporal coupling between the P3 and the response.

The P3 is therefore often investigated following stimuli to which subjects respond directly. While overt responses are not a necessary precondition for P3 elicitation, if overt responses do occur, they are typically aligned with the P3. Specifically, a frontal instance of the P3-family peaks slightly before the response, while the P3b typically peaks just at, or rapidly following it (Delorme, Sejnowski, & Makeig, 2007a; Makeig et al., 1999, 2004). However, the P3 is not a motor component. A P3 is found in response inhibition trials (Falkenstein, Hoormann, & Hohnsbein, 1999). Furthermore, direct comparisons between overt and covert tasks have demonstrated that the P3 is also observable in passive (task-free) paradigms (e.g. in response to incorrect sequence endings), with P3 amplitudes typically (but not always) smaller than in the presence of an active task (see Lang & Kotchoubey, 2002, and the references cited therein). In one study, a silent counting task even increased P3 amplitude (Salisbury, Rutherford, Shenton, & McCarley, 2001).

One reliable exception to the tight coupling between response timing and P3 latency is found when response selection is rendered complicated, for example by introducing incompatible stimulus–response mappings or complex motor actions (Verleger, 1997). Stressing speed over accuracy (Kutas, McCarthy, & Donchin, 1977) also dissociates RT and P3. Dissociations between RT and

P3 have inspired attempts to use the P3 as an indicator of stimulus evaluation time (rather than response selection time) and, accordingly, P3 latency as a measure of “mental chronometry” (McCarthy & Donchin, 1983). However, while Kutas and colleagues observed an RT/P3 dissociation when instructions emphasised speed over accuracy and a high RT/P3 correlation when accuracy was emphasised, other studies (Pfefferbaum, Ford, Johnson, Wenegrat, & Kopell, 1983) have reported exactly the opposite pattern (i.e. a low RT/P3 correlation under accuracy-emphasising instructions). On the basis of a comprehensive review of the P3 literature Verleger (1997; see also Verleger, 1988, 2010) argues against the stimulus evaluation view of the P3 by demonstrating that P3 latency has proven sensitive to a wide range of factors that also affect reaction times. P3/RT alignment holds as long as RTs in the fastest condition are brief (i.e. not drawn out by e.g. incompatible stimulus–response mappings). Verleger thus suggests that the P3 implements a linking between stimulus-induced and response-oriented processes.

Notably, in single-trial analyses of P3 data as visualised by ERP-images (Jung et al., 1999; see also Section 2.5 for a more detailed description of the ERPimage methodology) the P3 reliably shows up as RT-aligned (e.g., Chennu, Craston, Wyble, & Bowman, 2009; Johnson & Olshausen, 2005; Jung et al., 2001; Makeig et al., 1999, 2004; Marathe, Ries, & McDowell, 2013; O'Connell, Dockree, & Kelly, 2012; Townsend et al., 2001). We are unaware of even a single study showing RT-sorted ERPimages where a late centroparietal positivity was not found to be RT-aligned. We hypothesise that RT/P3 dissociations appear under two circumstances: either when selecting to respond is not immediately followed by a response because response selection and execution of responses is made difficult; or when the low signal-to-noise ratio of the EEG disallows a precise estimate of single-trial P3 latencies, for example, because wide RT variance leads to large search windows or because low confidence leads to low amplitudes.

In the language domain, researchers typically hope to avoid P3 “contamination” by asking subjects to delay response execution for some time after stimulus presentation. However, direct comparisons of immediate-response and delayed-response tasks demonstrate that, if at all, the P3 is slightly attenuated, but not abolished by response delay (Grent-t-Jong et al., 2011; Praamstra, Meyer, & Levelt, 1994; Smith, Barry, & Steiner, 2013). Phrased differently, in immediate-response tasks, the P3 follows the stimulus and is aligned to the response. In delayed-response tasks, by contrast, subjects are presented with sequences which sometimes contain a certain element (such as a target item) and, after each sequence, are asked to indicate via manual responses if the sequence did or did not contain an element of this class. In these studies, a P3 also follows the element licensing the *selection* of the response (i.e. the target), not the element licensing its *execution* (i.e. the response prompt). Thus, delayed-RT tasks do not fulfil the goal of removing response selection processes (as measured by the P3) from within the sequence. As will become clear below (see Sections 1.2 and 1.3), this observation is important for the design of the present study, which aimed to examine whether the P600 resembles the P3 in terms of being response-aligned.

1.2. The relation between the P600 and the P3

In their commentary on Coulson et al.'s (1998a) arguments in favour of the P600-as-P3 hypothesis, Osterhout and Hagoort (1999) noted: “[T]he actual testing of specific psycholinguistic models can profit from the existence of qualitatively distinct, language-relevant ERP effects, the P600/SPS not excluded [...] even though the actual cognitive and biological processes underlying these ERP effects remain obscure” (Osterhout & Hagoort, 1999, pp. 12–13). However, in attempting to move towards neurobiological models

of language (cf. Small, 2008), this is no longer a trivial assumption. To the contrary: the biological processes underlying language-related ERP effects become highly relevant. We thus argue that, for furthering our knowledge with respect to the neurobiology of language, the examination of the P600-as-P3 hypothesis is interesting not so much for questions of nomenclature (i.e. whether it is appropriate to label the P600 a P3) nor for questions of language-specificity versus domain-generality. Rather, if the P600 shows similar response properties to the P3, this would allow us to draw upon the considerable progress that has been made over the past decades in understanding the neurobiological basis of the P3 in order to illuminate the neural mechanisms of language processing. As we will discuss in more detail in Section 1.3, we view the LC/NE theory of the P3 as a particularly interesting approach in this regard. Thus, when referring to the “P600-as-P3” hypothesis (or, when appropriate, to the more specific “P600-as-LC/NE-P3” hypothesis) throughout this paper, we use this as a shorthand for the hypothesis that the P600 shares response properties/neurobiological underpinnings with the P3. Before describing the LC/NE account in a bit more detail, we will first present a very brief overview of prominent findings regarding the possible identity of the P600 and the P3.

As has been noted previously (e.g. Coulson et al., 1998a), including in the very first discussions of the P600 (Osterhout & Holcomb, 1992), the P600 and P3 resemble each other in general morphology and time course: both are late, positive components, prototypically with a centro-parietal maximum. They are also similar in terms of their antecedent conditions. A P600 often follows surprising, incongruent, intrusive words; often, such words are also task critical (e.g. in acceptability judgement tasks, as used for example by Osterhout & Holcomb, 1992). Consequently, from a domain-general perspective, it would not be unexpected to observe a P3 following such stimuli.

As discussed in Section 1.1, sentence processing experiments typically attempt to separate linguistic and task-related (P3) processes by delaying response execution until after the end of the sequence. However, the P3 is not elicited by overt responses. Rather, it indexes item classification and response selection (Verleger et al., 2005), and delayed-RT and immediate-RT iterations of the same paradigm elicit a nearly identical P3 (see above). This stands in contrast to other ERP components such as the CRN/ERN, which depend on overt motor responses.

The P600 appears in very similar contexts as the P3. Syntactic violations, by their very nature as violations, are salient and can be expected to elicit a P3. In line with this view, P600 amplitude is reduced when syntactic violations become common (Coulson et al., 1998a). When studies compare the same stimuli presented during explicit and passive tasks, the P600 is reliably larger when syntactic violations are task relevant, and may become small or absent when they are not (Hahne & Friederici, 2002; Haupt, Schlesewsky, Roehm, Friederici, & Bornkessel-Schlesewsky, 2008; Osterhout, Allen, McLaughlin, & Inoue, 2002; Osterhout, McKinnon, Bersick, & Corey, 1996). Furthermore, Hanulíková, van Alphen, van Goch, and Weber (2012) found that identical syntactic violations in Dutch only elicited a P600 when recorded by a native speaker of Dutch, but not when spoken by an L2-speaker with an obvious accent, thereby again supporting the idea that stimulus quality per se is not the most important factor with regard to the question of whether a P600 occurs or not.

This conclusion is further underscored by the observation that, when subjects do not attend to sentences that elicit a P600 when attended to, syntactic violations elicit early negative ERP components, but not necessarily a P600 (Batterink & Neville, 2013; Hasting & Kotz, 2008). While the N400, for example, is sometimes assumed to be a stable marker of automatic processing (Luck, Vogel, & Shapiro, 1996), the P600 is therefore labile under reduced

conscious awareness. This mirrors the dependence of the P3 on the subjective salience and significance of a stimulus (Nieuwenhuis, Aston-Jones, & Cohen, 2005; Spencer et al., 2001); components such as the MMN remain stable regardless of attention and awareness, but the P3 depends on subjective salience.

A major controversy then concerns whether the P600 is evoked only by specific structures (such as structural anomalies), unlike the exogenous P3, which depends not on inherent properties of the stimulus, but on its subjective significance. A large body of work argues for the reliance of the P600 on specifically structural violations and phenomena (Gouvea et al., 2010; Osterhout & Hagoort, 1999; for discussion and a different view, see also Coulson, King, & Kutas, 1998b; Coulson et al., 1998a). In many studies, a P600 follows only structural, but not, for example, semantic violations (e.g. Osterhout & Nicol, 1999; Osterhout et al., 2002), supporting its traditional interpretation as a specific index of structural processing. Other studies, however, show P600 effects following violations of thematic structure (for recent reviews, see Bornkessel-Schlesewsky & Schlesewsky, 2008; van de Meerendonk, Kolk, Chwilla, & Vissers, 2009), pragmatic manipulations (Burkhardt, 2007), purely physical manipulations such as visual degradation (van de Meerendonk, Chwilla, & Kolk, 2013), or following semantic anomalies, semantic judgement tasks or misspelt words (Fischler, Childers, Achariyapaopan, & Perry, 1985; Roehm, Bornkessel-Schlesewsky, Rösler, & Schlesewsky, 2007; Sanford, Leuthold, Bohan, & Sanford, 2011; van de Meerendonk, Indefrey, Chwilla, & Kolk, 2011; Vissers, Chwilla, & Kolk, 2006). For more than three decades, semantic violations have been found to induce strong P600 effects, both sentence-finally (Kutas & Hillyard, 1980, Fig. 1b and c) and in sentence-intermediate positions (Faustmann, Murdoch, Finnigan, & Copland, 2005; Hagoort, Wassenaar, & Brown, 2003; van Herten, Kolk, & Chwilla, 2005; even during passive processing of multi-sentence stories: Münte et al., 1998; Szewczyk & Schriefers, 2011). Though the affinity of the P600 for structural violations must be explained, it is clearly not specific to structural violations. However, the question remains why syntactic anomalies appear to evoke a P600 more readily than semantic ones.

As demonstrated by van de Meerendonk et al. (2010), strong, salient (“deeply implausible” in van de Meerendonk et al.’s terminology) semantic anomalies induce a P600 (following an N400), while more subtle (“mildly implausible”) anomalies only engender an N400. A similar dependence of the P600 on the intrusiveness and task-relevance of a semantic violation was also reported by Geyer, Holcomb, Kuperberg, and Perlmutter (2006) (for a discussion of these and further factors affecting the presence or absence of P600 effects to semantic anomalies, see Szewczyk & Schriefers, 2011). These findings corroborate Coulson et al.’s (1998a) suggestion that the stronger propensity of syntactic violations for eliciting P600 effects could be due to the more strongly categorical nature of syntactic violations as opposed to semantic anomalies. Accordingly, they predicted that semantic violations should also engender P600 effects when they are easy to classify as outrightly unacceptable – as is the case for intrusive, salient semantic anomalies. Similarly, a late positivity has been reported for semantically unexpected words in emotionally salient, but not neutral sentences (Moreno & Rivera, 2013). This observation converges with the P600-as-P3 approach, where the P600/P3 reflects the subjective significance of an item. Under this account, the late positivity is a measure of salience and thus becomes a gauge of the subjective significance of words.

Arguments based on scalp topography, source localisation and component additivity are inconclusive, since a reliable inverse model of ERP generation is not available. The P600 and P3 display similar topologies, but this does not necessarily imply neurophysiological equivalence. Additivity (i.e., the observation that combining a linguistic P600-eliciting and a non-linguistic P3-eliciting

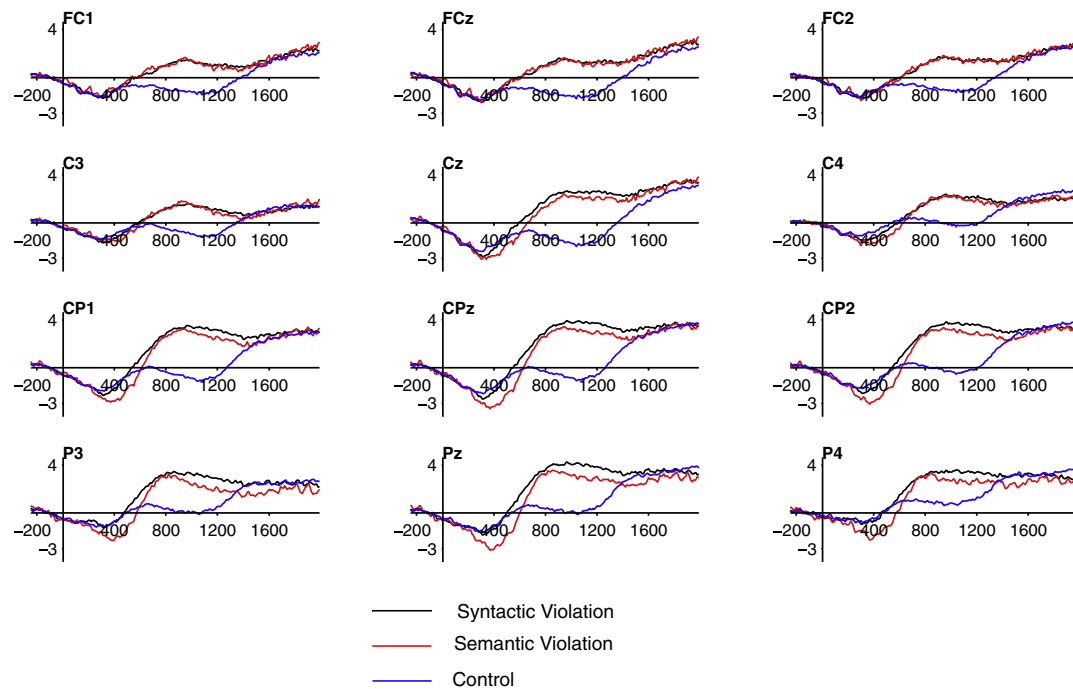


Fig. 1. Grand average ERPs for morphosyntactic violations (black trace), semantic violations (red trace) and control sentences (blue trace) at 12 representative electrodes. Positivity is plotted upwards.

feature leads to an ERP that resembles the linear summation of P600 and P3; see [Osterhout et al., 1996](#)) has been used to argue against the P600-as-P3 hypothesis, since the scalp effect of independent cortical generators is the linear sum of individual contributions. Conversely, in defence of the P600-as-P3 hypothesis, [Coulson et al. \(1998a\)](#) argued that P3 magnitude correlates with item salience and subjective categorisation confidence, and double violations are presumably more salient.

Further studies arguing against the P600-as-P3 perspective report that basal ganglia ([Frisch, Kotz, Cramon, & Friederici, 2003](#)) or Broca's area ([Wassenaar, Brown, & Hagoort, 2004](#)) lesions eliminate a linguistic P600, yet not an oddball P3 (though several studies also report a P600 after left-hemispheric or basal ganglia lesions; [Kielar, Meltzer-Asscher, & Thompson, 2012](#); [Kotz & Friederici, 2003](#), indicating that task peculiarities may be responsible rather than a specific role of the lesioned area in P600 generation). In these studies, linguistic but not oddball task performance was drastically impaired in the lesion group compared to controls, thus in fact strengthening the link between the P600 and behaviour, and thereby, the P3. The missing P600 here may simply reflect that participants were not able to reliably realise that an item should be categorised as ungrammatical.

1.3. The LC/NE-P3 theory

A recent account of the P3 side-steps many of these issues (e.g. co-localisation of P3 and P600 to common cortical or subcortical generators), while at the same time entailing a novel range of predictions under the assumption that it also applies to the P600. In contrast to models explaining ERP generation by the evoked synchronisation of independent cortical generators, [Nieuwenhuis et al. \(2005\)](#) connect the P3 to phasic activity of the brainstem Locus Coeruleus/LC ([Aston-Jones & Cohen, 2005](#); [Bouret & Sara, 2005](#)). They thus associate it with a neuromodulator system affecting multiple cortical sites with a distinct time course. The LC diffusely releases norepinephrine/NE, which facilitates general cortical state transitions, supporting cognitive reorientation (like

response execution or inhibition). The P3 is mostly insensitive to the sensory qualities of the stimulus and reflects contextually evoked subjective significance: surprising or expected, task relevant or intrusive stimuli may all result in a P3, since they all require cortical reorientation. Accordingly, the P3 has also been connected to the Ventral Attention Network ([Corbetta, Patel, & Shulman, 2008](#)), which governs effective stimulus-driven reorienting. This system is activated by stimuli such as task-critical targets, which, by their subjective importance, capture the subject's attention. This strict association between the timing of the P3 and that of overt behavioural responses is emphasised in the LC/NE-P3 theory, since this same alignment between overt, behavioural manifestations of reorientation mirrors that of LC neurons, which are known to be better aligned with response than with stimulus timing ([Rajkowski, 2004](#)).

The topography and latency of the P3 complex mirrors the distribution of ascending LC projections ([Nieuwenhuis et al., 2005](#)). According to the LC/NE theory of the P3, these correlations result from a causal relationship: the NE impulse from the LC both causes the synchronised depolarisation resulting in the scalp P3 as well as facilitating the behavioural response. Therefore, P3 and behaviour correlate on a single-trial level.

[Nieuwenhuis et al. \(2005\)](#) propose that, following the decision about stimulus significance (categorisation of the stimulus into a class of items requiring state transitions in light of the current strategy), an LC release of NE facilitates the selection of appropriate responses, regardless of the nature of the response (e.g. movements or memory updating). The P3's RT-alignment also results from a causal relationship: NE from the LC facilitates state shifts and causes the P3. We thus focus on the LC/NE theory of the P3 here since this account is not only neurobiologically explicit, but also, of the current P3 theories, it is the one that most directly predicts response-alignment.

1.3.1. The present study

In our view, the previous findings outlined in Section 1.2 are consistent with the P600 as a marker of subjective significance of

linguistic material, rather than of structural processing. Here, we put this hypothesis to a critical test by investigating if the late positivity following structurally deviant linguistic material shows the RT-alignment typical of the P3, as predicted by the P600-as-LC/NE-P3 hypothesis.

RT alignment is neither a necessary nor an obvious feature of theories assuming that the P600 reflects linguistic processing or other aspects of stimulus analysis. Post-hoc additions to such theories could explain RT alignment of the P600. However, as discussed in Section 1.1, the relationship between P3 latency and RT is reliable. A dissociation between P600 latency and RT would falsify critical predictions of the P600-as-P3 hypothesis. Previous research demonstrated RT alignment of the error-related negativity (Debener et al., 2005) and multiple members of the P3 family (Makeig et al., 2004), and onset alignment of N100/P100 (Jung et al., 2001). Cummings et al. (2006) found that a stimulus-interpretative component, the N400 (Kutas & Federmeier, 2011), is aligned to stimulus onset, not RT, thereby establishing that late, high-level components can be stimulus aligned.

Previous sentence processing experiments lack the required information for investigating RT alignment of components. Either no overt task was used, or the task was delayed relative to the critical stimulus. We are not aware of previous electrophysiological sentence processing studies in which participants judged linguistic deviancy as soon as they detected the error, allowing for a correlation of RT and P600 latency. The present study aimed to fill this gap.

Three alternative outcomes appear possible

- (a) The late positive component following deviant structures could appear aligned to stimulus onset, with no RT-locked positive peak observable;
- (b) two distinct positive components could appear, which are RT- and onset-locked, respectively;
- (c) only one, RT-aligned late positive component could appear.

Scenarios (a) and (b) would both falsify an essential prediction of the P600-as-LC/NE-P3 perspective, by providing evidence for a distinct P600 component. Scenario (c), by contrast, is predicted by the P600-as-P3 perspective, while models assigning the P600 a specific role in structural/combinatorial processing might require post hoc amendments to explain this scenario.

The present study aimed to test these hypotheses. Please note that, in line with recent calls for dissociating exploratory from confirmatory research (Wagenmakers, Wetzels, Borsboom, van der Maas, & Kievit, 2012), we pre-registered the experiment (German Clinical Trial Registry, ID: DRKS00004596), making our predictions and methods publicly available before data collection was initiated.

2. Materials and methods

2.1. Participants

Twenty monolingually raised native speakers of German (three men; mean age 24.75, range 21–42) participated in the experiment after giving written informed consent. Participants were right-handed, had good auditory acuity and normal or corrected-to-normal vision. All were students of the University of Mainz, receiving course credit for their participation.

2.2. Materials

Experimental stimuli were constructed by a strict scheme, resulting in sentences of the structure shown in example (1). Each

sentence consisted of a hyperonym and two potential hyponyms, always presented in that order. Only these three nouns and their determiners were varied across sentences. Control sentences (1a), of which subjects heard 150, contained a hyperonym and two hyponyms. Syntactic violation sentences (1b), of which subjects heard 110, consisted of a hyperonym and two of its hyponyms, one of which (balanced across 1st and 2nd positions) was preceded by an article not agreeing in grammatical gender with the hyponym. Agreement violations, including gender mismatches, have previously been found to elicit P600 effects (Hagoort & Brown, 1999; Molinaro, Barber, & Carreiras, 2011). Semantic violations (1c), of which subjects heard 40, consisted of a hyperonym, one of its hyponyms, and one noun phrase that had been exchanged with a noun phrase from another sentence. Semantic errors of this sort typically induce N400 effects (Kutas & Federmeier, 2011), sometimes followed by an additional P600 (e.g. Roehm et al., 2007; Sanford et al., 2011).

We used a higher number of sentences in the two conditions of primary interest – the control condition and the syntactic violation condition, where we expected to observe a P600 – than in typical studies of sentence processing in order to enable us to conduct single trial analyses. Because we were unable to produce 300 different hyponyms, many hyponyms were shared across sets. However, we ensured that no sentences were repeated verbatim, and neither condition (structural violation, semantic violation or correct) nor violation time point were predictable before the actual violation point/critical point (1st or 2nd hyponym for violation sentences, and 2nd hyponym for control sentences). Instead, morphosyntactic agreement of each hyponym with its determiner, and semantic fit of each hyponym with its hyperonym had to be computed.

- (1) Example sentence stimuli from the present study (critical word in bold)
- | | |
|-----|---|
| Con | a. Zur Kategorie Getränke gehören die Fanta und das Wasser.
To the category drinks belong the.fem Fanta.fem and the.neut water.neut
“Fanta and water belong to the category drinks.” |
| Syn | b. *Zur Kategorie Getränke gehören der Fanta und das Wasser.
To the category drinks belong the.masc Fanta.fem and the.neut water.neut |
| Sem | c. *Zur Kategorie Getränke gehören die Qualle und das Wasser.
To the category drinks belong the.fem jellyfish.fem and the.neut water.neut
“Jellyfish and water belong to the category drinks.” |

All comparisons were undertaken using identical words (i.e. in a different lexical set, “Qualle”/jellyfish occurred in a correct control condition). All 3 German genders appeared approximately equally often in both nouns and articles.

Sentences were recorded by a trained speaker and onset times of critical words extracted for EEG time-locking. Two randomised lists of 300 sentences (150 correct control sentences, 110 syntactic and 40 semantic violations) were constructed. In each list, 100 sentences contained unique lexical material, 100 hyponyms were used in two sentences of different conditions.

2.3. Procedure

Participants were seated in a soundproofed booth in front of an LCD monitor and listened to sentences presented via loudspeakers. Each trial began with the presentation of a neutral smiley at the centre of the white background. After 100 ms, sentence presentation

began. Participants were instructed to attend to the sentence and press either a left- or a right-hand button as soon as they realised whether the sentence was correct or deviant in form or content. Following the button press, the neutral smiley was replaced with a feedback smiley indicating either a correct (smiling smiley) or missed ($RT > 3$ s)/incorrect (frowning smiley) answer. After 1000 ms, the next trial began. Assignment of left and right buttons to incorrect and correct was counterbalanced across participants.

After each 20-trial block, feedback screens presented that block's mean RT and error rate. Experimental sessions, including electrode application, lasted 1.5–2 h.

2.4. EEG recording and analysis

The EEG was recorded with 32 Ag/AgCl electrodes and a left mastoid reference, using a BrainProducts BrainAmp (Brain Products GmbH, Gilching, Germany) and processed in EEGLAB/ERPLAB (Delorme & Makeig, 2004; Lopez-Calderon & Luck, 2014). Data sets were bandpass filtered at 0.1–40 Hz, average re-referenced (Dien, 1998) and epoched around correct responses. Incorrectly answered or artifactual epochs (identified by kurtosis statistics; Delorme et al., 2007a) were excluded. After Extended Infomax ICA decomposition (Makeig et al., 1999), blink/vertical eye movement components were automatically rejected (Viola et al., 2009), although this had little effect on the measures at electrode PZ.

For each individual participant, single-trial difference waves (Bishop & Hardiman, 2010) at electrode PZ were created by subtracting the mean (onset-locked) ERP of control sentence hyponyms from each individual semantic or morphosyntactic violation trial. Note that even though control sentences were also responded to, there, participants had to withhold responses until the second noun and therefore, only 50% of control hyponyms were immediately followed by a response.

As noted in Footnote 1, all scripts for data analysis have been uploaded to a public repository and can be accessed at <https://github.com/jona-sassenhagen/Charybdis>.

2.5. ERP and ERPimage analysis

ERPs were plotted using ERPLAB. The difference between mean ERP amplitude in syntactic and semantic violation trials in the P600 time window (500–1000 ms) at electrode PZ was submitted to a paired, two-tailed *t*-test, which indicated that mean amplitude was higher (i.e. more positive) for syntactic violations ($t(19) = 3$; $p = 0.006$; 95% CI = 0.3–1.5).

All further analyses were conducted on difference trials at electrode PZ.

RT- sorted ERPimages provide a straightforward method for investigating RT alignment (Jung et al., 2001). In ERPimages, multiple event-locked EEG epochs (trials) are stacked horizontally as colour-coded lines, showing time on the *x* axis and trial number on the *y* axis, with colour indicating time-trial point potential. After visual smoothing, this provides the same information as an ERP: horizontal red lines, indicating potential mean-positive windows, correlate with positive ERP peaks, blue lines correlate with negative peaks. ERPimages can be sorted by various measures, especially event latencies. Time-locking to stimulus onset and sorting by RT, stimulus-aligned components appear as horizontal lines parallel to onset, RT-aligned components diagonal/sigmoidal, parallel to RT.

Since no single standard method for quantifying RT alignment has been established, we employed three different methods that have all been previously shown to indicate RT-alignment of the P3: latency estimation of RT bin, Woody filter estimation of single-trial latencies allowing single-trial correlations, and inter-trial phase coherence of RT- versus onset-aligned data.

2.5.1. RT binning

This conceptually simple, transparent and popular method (Marathe et al., 2013; Poli, Cinel, Citi, & Sepulveda, 2010; Roth, Ford, & Kopell, 1978) has repeatedly shown P3 latency to correlate with RT. It comprises binning individual subjects' trials by RT quartile, estimating the latency of ERP components per bin, and analysing if latency increases with bin rank.

Following standard procedures (Kiesel, Miller, Jolicoeur, & Brisson, 2008; Luck, 2005; Ulrich & Miller, 2001), we excluded the top and bottom 2.5% of trials for each subject, binned by individual subject RT quartile, set all negative values to zero to avoid contributions from the N400, constructed jackknife averages and estimated the 33% fractional latency of the area under the positive curve. If the P600 is RT aligned, slower bins must show later latency estimates. A repeated-measures one-way ANOVA with the factor RT quartile was applied to test the statistical reliability of this effect. The outcome was corrected for the jackknife procedure (Kiesel et al., 2008).

2.5.2. Woody filtering

Kutas et al. (1977) applied a Woody filter (Woody, 1967) to identify single-trial P3 latencies and found a strong correlation ($r = 0.42$ – 0.66) with RT. We implemented a Woody filter as follows: We calculated a subject mean ERP for syntactic violation difference trials with RTs between 500 and 1250 ms. We then established the time lag of the best correlation between this ERP and each single trial of the same subject in a window from 500 to 1500 ms after stimulus onset. For 100 iterations, a new template ERP was calculated by shifting each trial by the identified lag, and the best correlation between the template and individual single trials was computed. The time point of best correlation between single trials and the final template iteration was taken as the latency of the late positivity.

We then calculated the skipped Pearson's correlation coefficient (Rousselet & Pernet, 2012) between single-trial RTs and positive component latency for individual subjects. Then, the same procedure was repeated for the late positivity and the N400 (time window: 0–550 ms) for semantic violations.

Problematically, we found that the *r* obtained from this measure greatly depended on the precise analysis parameters such as window onset and length.

2.5.3. RT- versus onset-aligned ITC

Inter-trial phase coherence (ITC; Delorme, Westerfield, & Makeig, 2007b) is a measure of cross-trial phase consistency of EEG oscillations. Comparing the same single-trial data under two different temporal alignments shows to which time point event-related perturbations are better aligned. ITC is calculated via wavelet decomposition of single trials and the computation of phase consistency per frequency and time point across individual trials.

A frontal P3 has been found to show higher phase consistency when trials were aligned to RT than to stimulus onset, indicating RT alignment. We calculated the time and frequency mean ITC from 0.5 to 8 Hz for each subject, separately for RT- and onset-aligned trials, in a 50 ms window focused on the positive peak (EEGLAB function `newtimef.m`, wavelet decomposition of data from electrode Pz, minimum 2 cycles, 4 s pre-stimulus single-trial baseline).

3. Results

Participants' overall accuracy on the judgment task was good (mean error rate: 11%; average RT for semantic violations: 831 ms, for morphosyntactic violations: 844 ms).

Fig. 1 shows ERPs to semantic and syntactic violations and control conditions. For semantic violations, a vertex-negative component peaked at around 450 ms, followed by a broad vertex-positive wave. Syntactic violations showed a similar late positivity, which was slightly more pronounced than that for semantic violations (paired *t*-test for amplitude differences between violation and control conditions at electrode PZ: $t(19) = 3$; $p = 0.006$; 95% CI = 0.3–1.5).

Fig. 2A shows RT-sorted violation minus control difference ERP-images of all participants' single trial EEG at PZ, aligned both to the onset of words inducing a morphosyntactic violation, and to RT, and the corresponding ERP. Onset-aligned ERP images (150-epoch Gaussian smoothing) revealed an onset-aligned P600 with a broad, flat morphology, whereas in RT-aligned ERPs, the component peaked sharply, corresponding to a focused positive component in the RT-locked ERP image. Semantic violation difference ERP images (see Fig. 2B) reveal a similar RT-aligned late positivity and a stimulus-aligned N400.

To quantify onset and reaction time locking, we employed three measures: RT bin peak latencies, Woody filter estimates of

component latency, and response- versus phase-locked ITC. For the syntactic violation condition, bin latency strictly increased with bin RT and RT bins were unlikely to reflect activity with identical latency (corrected $F(3,76) = 28$; $p < 0.0001$). Bin latency monotonically rose with bin RT (mean 33% fractional area latency and mean bin RT for fastest to slowest bin: 770/606, 854/760, 926/920 and 1037/1190 ms; Spearman's $\rho = 1$). RT quartile-binned ERP latencies also correlated with mean bin RT for semantic violation trials ($\rho = 1$).

Woody filter-estimated single-trial latencies of the late positivity following syntactic violations correlated strongly with single-trial RT (95% CI: .5, .73), but the N400 following semantic violations did not (95% CI: $-.1, .22$).

During the P600 peak window, phase locking of low-frequency activity (as measured by ITC) was greater for RT-aligned than for onset-aligned trials (95% CI: 5.4–11% greater ITC for RT-aligned trials).

Parameter estimates for the Woody filter and ITC analyses are summarised in Table 1.

4. Discussion

The present study used single-trial EEG analyses to distinguish response – from stimulus-aligned effects in a linguistic deviancy detection task including button presses directly following critical parts of the sentence. The late positive EEG deflection following linguistically deviant material was strictly RT aligned, with no distinct, second positive peak aligned to stimulus onset. The N400 following semantic deviations behaved like an exogenous component in that it was stimulus – rather than response-aligned (compatible with Cummings et al., 2006). These results confirm an important prediction of the P600-as-LC/NE-P3 perspective. A dissociation between RT and P600 would have falsified this theory; the positive finding allows for a neurophysiological grounding of the P600 by association with the LC/NE system (Nieuwenhuis et al., 2005).

4.1. Limitations of the current approach and consequences for future research

4.1.1. Possible caveats due to an “unnatural” paradigm?

It could be argued that the repetitive nature of our stimuli and our explicit task caused the sentences to be perceived in a more task-heavy processing mode, causing the appearance of a P3-like component instead of the components expected for more naturalistic stimuli. However, to complete the task successfully, subjects necessarily processed the stimuli semantically and syntactically up to a depth sufficient for anomaly detection. Thus, if the (only) observed positivity is a P3, the question then becomes: where is the P600? If the present late positivity is a P3, the lack of a distinct P600 entails that there is no P600 as a general, necessary consequence of syntactic processing, or at the very least that it depends on specific (as of yet unspecified) aspects of the task. In either case, a model of the P600 as natural correlate of automatic syntactic processing must be amended.

In addition, the assumption that the present paradigm only elicited a P3 but no P600 is at odds with results demonstrating that the P600, in fact, has a stronger propensity to appear in task-relevant contexts than when task relevance and syntactic manipulation status do not coincide. As noted in the introduction section, the P600 – following both syntactic and semantic anomalies – is enhanced by more explicit tasks (Hahne & Friederici, 2002; Haupt et al., 2008; Osterhout et al., 1996, 2002). It is greatly attenuated and often absent (Batterink & Neville, 2013; Hasting & Kotz, 2008; Royle, Drury, & Steinhauer, 2013) when subjects do not consciously attend to grammatical violations – in contrast to syntax-sensitive

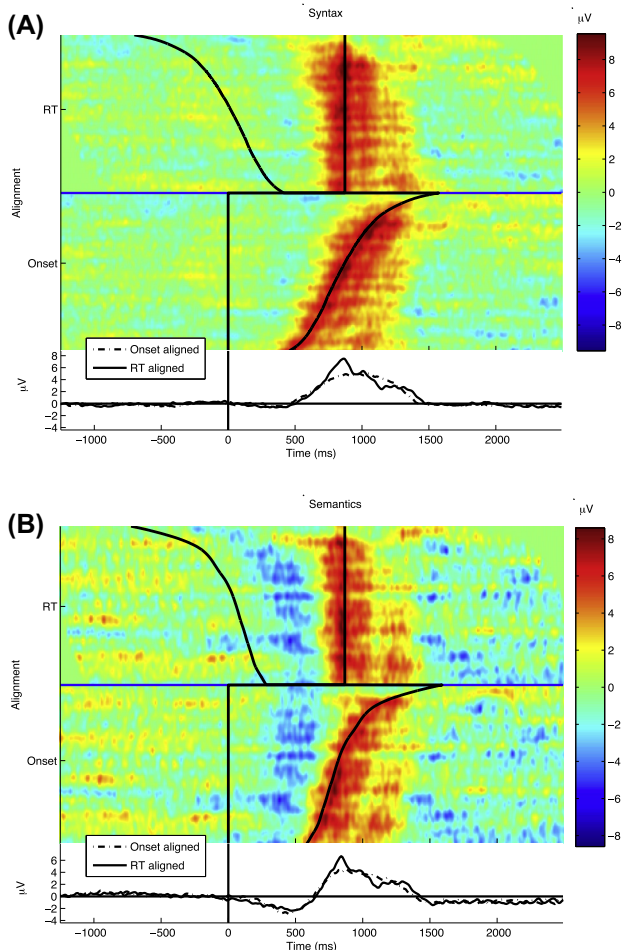


Fig. 2. RT-sorted ERP images of difference waves at PZ (violation minus control difference, Gaussian visual smoothing). (A) morphosyntactic violations. (B) semantic violations. In each case, the top panel shows data time-locked to responses (vertical line), while the bottom panel shows data time-locked to stimulus onset (vertical line). In the top panels, the curved line corresponds to stimulus onset, while in the bottom panels, the curved line corresponds to responses (RT). As is clearly apparent from the figure, the late positivity is response-locked in all cases (A and B). The N400 for semantic violations (B), by contrast, shows characteristics of stimulus-locking (i.e. alignment with the curved line in the top panel and with the vertical line in the bottom panel). For further details on the interpretation of ERP images, please see the main text.

Table 1

Summary of the parameter estimates for RT-locking of the ERP components observed in the present study for the Woody filter and ITC difference methods. For the results of the RT bin analyses, please see the main text. Point estimates for intervals excluding 0 (entailing significant tests) are marked in bold.

Measure	Contrast	95% CI lower bound	Mean	95% CI upper bound
Method: Woody Filter estimated latency and RT correlation Unit: Correlation coefficient	Syntactic violations: late positivity	0.5	0.63	0.73
	Semantic violations: late positivity	0.42	0.59	0.72
	Semantic violations: N400	−0.02	0.09	0.12
Method: Low-frequency phase coherence Unit: ITC difference (ratio)	Syntactic violations: RT- versus onset-aligned	0.05	0.08	0.12
	Semantic violations: RT- versus onset-aligned	0.06	0.09	0.12

negativities, which often remain rather unaffected by task (e.g. Haupt et al., 2008).

It also appears highly unlikely that the use of an immediate-response paradigm led to a higher likelihood for a P3 in this study as opposed to previous sentence processing experiments employing similar violation paradigms and delayed reaction. It has been established that the P3 follows the event affording decision making and response selection, not response execution. A direct comparison of immediate and delayed response tasks (e.g. Grent-'t-Jong et al., 2011; Praamstra et al., 1994) reveals that a P3 is always seen on the critical stimulus itself, whether it is immediately followed by a response or not. In other words: the P3 does not “wait for the ‘go’ signal”. In accordance with these findings from non-linguistic paradigms, a P3 is expected following task-relevant violations in typical (delayed-response) EEG sentence processing experiments just as for the present immediate-response paradigm.

Finally, it may be questioned if passive perception and comprehension is indeed the more “natural” mode of language processing, as opposed to “preparation for situated action” (Barsalou, 1999).

In summary, when the present study is considered in light of the full range of existing data, there is no principled reason to assume that the paradigm employed here should have been more susceptible to eliciting a P3 effect than previous violation studies on sentence processing. The fact that the only positivity following the processing of structural information in our study is RT-aligned thus has implications for our understanding of the P600.

4.1.2. Predictions for future, more naturalistic studies

The present, highly regulated paradigm has provided some new insights into the response properties of the P600 by demonstrating its response alignment. In future work, more naturalistic paradigms could be employed to test other predictions of the P600-as-LC/NE-P3 hypothesis. These include the testable prediction that other factors that covary with activation of the LC system such as pupil dilation, heart rate increases and skin conductance responses (Nieuwenhuis et al., 2005) should react to syntactic deviances the same way as the late positivity. Moreover, late positivity effects should be modulated by these independent physiological criteria. Specifically, we speculate that individual differences in the presence or absence of late positivity effects in a particular language processing paradigm (e.g. Bornkessel, Fiebach, & Friederici, 2004; Nakano, Saron, & Swaab, 2010; Nieuwland & Van Berkum, 2008; Osterhout, 1997; Roehm et al., 2007) may be explainable in terms of such physiological parameters, reflecting the subjective salience of a stimulus to a participant rather than qualitatively different analysis strategies (e.g. in terms of semantic versus syntactic analysis).

4.2. Electrophysiological correlates of analysing linguistic information

The alignment of the P600 to RT is not directly predicted by accounts assuming that the P600 reflects a process related to the (re)structuring of the linguistic input. In single trials, the behavioural responses are aligned to a point in time that falls under

the P600 curve (cf. the red amplitude markers in the ERPimages and the correlation between RT and peak P600 latency).

For a process-based account (in terms of more effortful structural analysis, reanalysis etc.), this entails that RT correlates with a specific time point within the overall process. How such a point might be defined is unclear. Instead, reanalysis- or repair-based interpretation of the P600 imply that the behavioural response correlates – at least to a certain degree – with the endpoint of the reanalysis/repair process, which should be reflected in P600 offset (i.e. a point that is no longer under the P600 curve). Since linguistic analysis still needs to be followed by response selection/motor disinhibition processes varying in length, strong RT correlations are not expected (cf., for example, speed-accuracy tradeoff effects in RT measures, which show that the reaction is, to some degree, independent of critical stimulus properties).

This argument concerns all approaches according to which the P600 reflects the (re)structuring or repair of linguistic input, independent of their specific interpretation of the types of processes involved (e.g. “late syntactic processes”, Friederici (2011, p. 1377); an “index for structural processing”, Kos, Vosse, van den Brink, & Hagoort (2010, p. 1); “attempts to create or repair syntactic relations”, Gouvea et al. (2010, p. 32); or “establishing a representation of what the speaker wants to convey”, Brouwer et al. (2012, p. 136)). We do not suggest that such accounts cannot explain P600 response alignment. However, whereas the P600-as-LC/NE-P3 account critically entails this finding, other models require additional assumptions to integrate it.

By contrast, existing sentence processing accounts which associate the P600 with the P3, such as the Monitoring Theory (e.g. van de Meerendonk et al., 2010; van Herten et al., 2005; Vissers et al., 2008), can account for the present results insofar as the P3 is known to be response-aligned (see Section 1.1), though the strength of that prediction will vary depending on the underlying model of the P3 that is assumed. The Monitoring Theory and the P600-as-LC/NE-P3 hypothesis both account for the present findings, in which we observed late positivity effects to ungrammatical – and hence unexpected – sentence continuations equally well. However, as the Monitoring Theory focuses particularly on unexpectedness as the primary antecedent of the P600/P3, the two approaches can be used to generate differing, testable predictions for future research. In particular, the P600-as-LC/NE-P3 hypothesis predicts that late positive ERP effects correlate with factors affecting the LC/NE system (e.g. heart rate, pupil dilation, see Section 4.1.2) should also be observable to *expected* stimuli that are rendered salient by some other property. An initial indication that this prediction may indeed be borne out is provided by the finding of late positivity effects in response to emotion words. This effect is largest for words with a negative emotional valency and is further modulated by task-relevance of the emotional content (e.g. Holt, Lynn, & Kuperberg, 2009; Kanske & Kotz, 2007; Kiehl, Hare, McDonald, & Brink, 1999). The negative-positive distinction is in accordance with observation that threatening stimuli show a particularly high inherent salience. From our understanding of the Monitoring Theory, this account would not directly predict late positivity effects to stimuli that are not unexpected, though it

may be possible to integrate such findings by assuming that inherently salient stimuli trigger monitoring processes.

From the perspective of the P600-as-LC/NE-P3 hypothesis, a challenge for future research will lie in the more precise characterisation of stimulus salience and, hence, subjective or motivational significance. We have proposed that late, language-related positivities can be interpreted primarily as a marker of subjective significance, which may come from an ill contextual fit sufficient to disrupt analysis, from task target status, or from the emotional value of a word. Beyond the complications arising from the complex interaction of these multiple factors, an operationalisation of subjective significance is rendered difficult by its inherently subjective nature: a stimulus may be more significant to some participants in a study than to others. Thus, at a first glance, the interpretation that late positive components in language processing simply measure subjective salience (e.g. of violations) may seem circular and unfalsifiable in itself. However, the theory entails a range of falsifiable predictions such as correlations with galvanic skin responses, heart rate and other physiological markers of LC/NE-system activation (see Section 4.1.2). In the present study, we tested response alignment as the simplest of these predictions. In future work, we aim to examine the effects of physiological parameters. By using these as proxies for ‘subjective significance’ and, thereby, as parametric predictors for late positivity responses, we aim to take the first step towards a more formalised and operationalisable definition of subjective significance of linguistic stimuli and the mechanisms involved in processing them.

Furthermore, while we have focused particularly on the neurophysiological aspects of reorientation as resulting from NE release in this initial investigation, it is clear that future work will need to spell out in more detail how cognitive reorientation translates into language processing mechanisms. *Sara and Bouret (2012)* describe the reorientation process as analogous to the “truncated conditioned reflex” (*Kupalov, 1961*), a conditioned reflex which manifests itself in changes in the functional states of the brain rather than in external behaviour. Essentially, this change can be viewed as an “increase in cortical arousal, attention, and expectancy” (*Sara & Bouret, 2012, p.133*). It facilitates memory retrieval as well as perceptual shifts when viewing ambiguous stimuli such as the Necker cube and may serve a resetting function to allow for changes to the focus of attention. As these functional properties help the organism to deal with unexpected input, e.g. by allowing for shifts of attention to the unexpected input item, this provides an interesting potential link to the cognitive assumptions of the Monitoring Theory (cf. *van de Meerendonk, Rueschemeyer, & Kolk, 2013*): the P3 as a marker of a shift of attention (see also Section 1.3), resulting from the saliency of an item, for example due to its unexpectedness, but also e.g. to emotional or degraded items, or expected, behaviourally critical events. Clearly, an important objective for future research is to investigate in detail the relation between these relatively general cognitive correlates of reorienting and mechanistic accounts of language processing. As already discussed above, we believe that situations involving expected, but subjectively significant stimuli in particular may help to provide important new insights on the precise mechanisms by which cortical reorientation induced by NE release relates to language processing.

In summary, we argue that, to explain the distribution of late positivities related to linguistic processing, nothing needs to be stipulated beyond the established understanding of the P3. Items that are particularly ill-fitting can be expected to disrupt analysis and evoke a positivity as a result of their high salience, as do items that belong to the category of task-critical events. Indeed, typical P600-inducing paradigms usually involve a much smaller set of task-critical events than typical N400-inducing paradigms: in the former, sentences are violations in virtue of being iterations of a

certain category (e.g. a specific incongruence of functional elements such as agreement or tense markers), while in the latter, violations arise in virtue of one token being incompatible in its inherent meaning with surrounding tokens. In addition, semantic anomalies may be less categorical in that, unless they are deeply implausible, they are less likely to be classified as outright violations (*Coulson et al., 1998a*). Accordingly, it appears reasonable that P600 effects are somewhat more likely to occur in response to (morpho-)syntactic as opposed to semantic violations (but see Section 1.2 for a discussion of P600 effects elicited by semantic incongruities).

5. Conclusions

In testing a critical entailment of the P600-as-LC/NE-P3 theory, we found that the late positivity following morphosyntactic violations behaved like a P3 in being response-aligned. Even though subjects successfully processed semantic content and syntactic structure, no distinct, stimulus-locked late positivity was observed. This result is predicted by all accounts which subscribe to the P600-as-P3 assumption, but requires additional post hoc assumptions for typical interpretations of the P600 as a distinct component reflecting the analysis, reanalysis or repair of linguistic input. While these results do not prove the P600-as-P3 hypothesis, they confirm a necessary entailment of this theory (particularly of the stronger, P600-as-LC/NE-P3 hypothesis), and any other finding would have strongly supported the hypothesis of a distinct P600 component.

Furthermore, we have demonstrated the feasibility of single-trial analysis techniques informed by immediate behavioural responses during stimulus presentation. Our findings show that single-trial analyses of sentence processing data can be used to inform models of the neurobiology of language.

Lastly, we would like to reiterate a point previously made by *Coulson et al. (1998a)*. Understanding the P600 as a type of P3 (i.e. as being traceable to the same underlying neurobiological system) does not automatically devalue it as a tool for the investigation of the neural substrates of language processing. If our interpretation of the late positivity in sentence processing experiments as an LC/NE-P3 is correct, this component marks a point in time where a linguistic entity has achieved subjective significance and some form of adaption process is underway. Its amplitude marks the degree to and reliability with which this stimulus class is significant. It thereby provides a gradient (though indirect, relative) measure for the time course of certain processes. It also reflects categorisation and recognition processes, marking the degree to which the stimulus category is consciously perceived as belonging to a certain class, such as experimental condition type from which tokens are repeatedly presented. While it may become necessary to give up on the idea of a specific EEG index of structural or combinatorial processing, a reliable measure for these fundamental mechanisms and how they contribute to language processing may be won instead.

Acknowledgments

Parts of the research reported here were supported by the German Research Foundation (BO 2471/3-2), the LOEWE programme of the German state of Hesse and by the University of Mainz as part of the research initiative “Pro Geisteswissenschaften”. We would like to thank Karin Hollerbach, Miriam Burk, Alexander Dröge and Phillip Alday for help with the stimulus materials and Brita Rietdorf for the data acquisition. We are also grateful to Herman Kolk and several anonymous reviewers for valuable comments on previous versions of the manuscript.

References

- Aston-Jones, G., & Cohen, J. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience*, 28, 403–450.
- Barsalou, L. W. (1999). Language comprehension: Archival memory or preparation for situated action? *Discourse Processes*, 28(1), 61–80.
- Batterink, L., & Neville, H. J. (2013). The human brain processes syntax in the absence of conscious awareness. *The Journal of Neuroscience*, 33(19), 8528–8533.
- Bishop, D., & Hardiman, M. (2010). Measurement of mismatch negativity in individuals: A study using single-trial analysis. *Psychophysiology*, 47(4), 697–705.
- Bornkessel-Schlesewsky, I., & Schlewsky, M. (2008). An alternative perspective on “semantic P600” effects in language comprehension. *Brain Research Reviews*, 59(1), 55–73.
- Bornkessel-Schlesewsky, I., Kretschmar, F., Tune, S., Wang, L., Genç, S., Philipp, M., et al. (2011). Think globally: Cross-linguistic variation in electrophysiological activity during sentence comprehension. *Brain and Language*, 117(3), 133–152.
- Bornkessel, I., Fiebach, C. J., & Friederici, A. D. (2004). On the cost of syntactic ambiguity in human language comprehension: An individual differences approach. *Cognitive Brain Research*, 21, 11–21.
- Bouret, S., & Sara, S. (2005). Network reset: A simplified overarching theory of locus coeruleus noradrenergic function. *Trends in Neurosciences*, 28(11), 574–582.
- Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about semantic illusions: Rethinking the functional role of the P600 in language comprehension. *Brain Research*, 1446, 127–143.
- Burkhardt, P. (2007). The P600 reflects cost of new information in discourse memory. *NeuroReport*, 18(17), 51–54.
- Chennu, S., Craston, P., Wyble, B., & Bowman, H. (2009). Attention increases the temporal precision of conscious perception: Verifying the neural-ST2 model. *PLoS Computational Biology*, 5(11), e1000576. <http://dx.doi.org/10.1371/journal.pcbi.1000576>.
- Corbetta, M., Patel, G., & Shulman, G. L. (2008). The reorienting system of the human brain: From environment to theory of mind. *Neuron*, 58(3), 306–324.
- Coulson, S., King, J., & Kutas, M. (1998a). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes*, 13(1), 21–58.
- Coulson, S., King, J., & Kutas, M. (1998b). ERPs and domain specificity: Beating a straw horse. *Language and Cognitive Processes*, 13(6), 653–672.
- Cummings, A., Ceponiene, R., Koyama, A., Saygin, A., Townsend, J., & Dick, F. (2006). Auditory semantic networks for words and natural sounds. *Brain Research*, 1115(1), 92–107.
- Debener, S., Ullsperger, M., Siegel, M., Fiehler, Cramon, K., Cramon von, D., et al. (2005). Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identifies the dynamics of performance monitoring. *Journal of Neuroscience*, 25(50), 11730–11737.
- Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1), 9–21.
- Delorme, A., Sejnowski, T., & Makeig, S. (2007a). Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *NeuroImage*, 34(4), 1443–1449.
- Delorme, A., Westerfield, M., & Makeig, S. (2007b). Medial prefrontal theta bursts precede rapid motor responses during visual selective attention. *Journal of Neuroscience*, 27(44), 11949–11959.
- Dien, J. (1998). Issues in the application of the average reference: Review, critiques, and recommendations. *Behavior Research Methods, Instruments, & Computers*, 30(1), 34–43.
- Donchin, E. (1981). Presidential address, 1980. Surprise! Surprise? *Psychophysiology*, 18(5), 493–513.
- Donchin, E. E., & Coles, M. G. H. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, 11(03), 357–374.
- Duncan-Johnson, C. C., & Donchin, E. (1977). On quantifying surprise: The variation of event-related potentials with subjective probability. *Psychophysiology*, 14, 456–467.
- Falkenstein, M., Hoormann, J., & Hohnsbein, J. (1999). ERP components in Go/Nogo tasks and their relation to inhibition. *Acta Psychologica*, 101(2–3), 267–291.
- Faustmann, A. A., Murdoch, B. E. B., Finnigan, S. P. S., & Copland, D. A. D. (2005). Event-related brain potentials elicited by semantic and syntactic anomalies during auditory sentence processing. *Journal of the American Academy of Audiology*, 16(9), 708–725.
- Fischler, I., Childers, D. G., Acharyapopan, T., & Perry, N. W. (1985). Brain potentials during sentence verification: Automatic aspects of comprehension. *Biological Psychology*, 21(2), 83–105.
- Friederici, A. (2011). The brain basis of language processing: From structure to function. *Physiological Reviews*, 91(4), 1357–1392.
- Frisch, S., Kotz, S. A., Cramon, D., & Friederici, A. (2003). Why the P600 is not just a P300: The role of the basal ganglia. *Clinical Neurophysiology*, 114, 336–340.
- Geyer, A., Holcomb, P., Kuperberg, G. R., Perlmutter, N. (2006). Plausibility and sentence comprehension. An ERP study. *Cogn. Neurosci. Suppl.*, Abstract.
- Gouvea, A. C., Phillips, C., Kazanina, N., & Poeppel, D. (2010). The linguistic processes underlying the P600. *Language and Cognitive Processes*, 25(2), 149–188.
- Gray, H. M., Ambady, N., Lowenthal, W. T., & Deldin, P. (2004). P300 as an index of attention to self-relevant stimuli. *Journal of Experimental Social Psychology*, 40(2), 216–224.
- Grent-t-Jong, T., Boehler, C. N., Kenemans, J. L., & Woldorff, M. G. (2011). Differential functional roles of slow-wave and oscillatory-alpha activity in visual sensory cortex during anticipatory visual-spatial attention. *Cerebral Cortex*, 21(10), 2204–2216. <http://dx.doi.org/10.1093/cercor/bhq279>.
- Hagoort, P., & Brown, C. (1999). Gender electrified: ERP evidence on the syntactic nature of gender processing. *Journal of Psycholinguistic Research*, 28(6), 715–728.
- Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, 8(4), 439–483.
- Hagoort, P., Wassenaar, M., & Brown, C. M. (2003). Syntax-related ERP-effects in Dutch. *Cognitive Brain Research*, 16(1), 38–50.
- Hahne, A., & Friederici, A. D. (2002). Differential task effects on semantic and syntactic processes as revealed by ERPs. *Cognitive Brain Research*, 13(3), 339–356.
- Hanulíková, A., van Alphen, P. M., van Goch, M. M., & Weber, A. (2012). When one person's mistake is another's standard usage: The effects of foreign accent on syntactic processing. *Journal of Cognitive Neuroscience*, 24, 878–887.
- Hasting, A. S. A., & Kotz, S. A. S. (2008). Speeding up syntax: On the relative timing and automaticity of local phrase structure and morphosyntactic processing as reflected in event-related brain potentials. *Journal of Cognitive Neuroscience*, 20(7), 1207–1219. <http://dx.doi.org/10.1162/jocn.2008.20083>.
- Haupt, F. S., Schlewsky, M., Roehm, D., Friederici, A. F., & Bornkessel-Schlesewsky, I. (2008). The status of subject-object reanalysis in the language comprehension architecture. *Journal of Memory and Language*, 59, 54–96.
- Holt, D. J., Lynn, S. K., & Kuperberg, G. R. (2009). Neurophysiological correlates of comprehending emotional meaning in context. *Journal of Cognitive Neuroscience*, 21(11), 2245–2262.
- Johnson, J. S., & Olshausen, B. A. (2005). The earliest EEG signatures of object recognition in a cued-target task are postsensory. *Journal of Vision*, 5(4), 299–312. [10:1167/5.4.2](http://dx.doi.org/10.1167/5.4.2).
- Jung, T.-P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., & Sejnowski, T. J. (1999). Analyzing and visualizing single-trial event-related potentials. *Advances in Neural Information Processing Systems*, 11, 118–124.
- Jung, T.-P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., & Sejnowski, T. J. (2001). Analysis and visualization of single-trial event-related potentials. *Human Brain Mapping*, 14(3), 166–185.
- Kaan, E. (2007). Event-related potentials and language processing: A brief overview. *Language and Linguistics Compass*, 1(6), 571–591.
- Kanske, P., & Kotz, S. A. (2007). Concreteness in emotional words: ERP evidence from a hemifield study. *Brain Research*, 1148, 138–148. <http://dx.doi.org/10.1016/j.brainres.2007.02.044>.
- Kiehl, K. A., Hare, R. D., McDonald, J. J., & Brink, J. (1999). Semantic and affective processing in psychopaths: An event-related potential (ERP) study. *Psychophysiology*, 36(6), 765–774.
- Kielar, A., Meltzer-Asscher, A., & Thompson, C. K. (2012). Electrophysiological responses to argument structure violations in healthy adults and individuals with agrammatic aphasia. *Neuropsychologia*, 50(14), 3320–3337.
- Kiesel, A., Miller, J., Jolicoeur, P., & Brisson, B. (2008). Measurement of ERP latency differences: A comparison of single-participant and jackknife-based scoring methods. *Psychophysiology*, 45(2), 250–274.
- Kos, M., Vosse, T., van den Brink, D., & Hagoort, P. (2010). About edible restaurants: conflicts between syntax and semantics as revealed by ERPs. *Frontiers in Psychology*, 1, 222. <http://dx.doi.org/10.3389/fpsyg.2010.00222>.
- Kotz, S. A., & Friederici, A. D. (2003). Electrophysiology of normal and pathological language processing. *Journal of Neurophysiology*, 16(1), 43–58.
- Kupalov, P. S. (1961). Some normal and pathological properties of nervous processes in the brain. *Annals of the New York Academy of Sciences*, 92, 1046–1053.
- Kutas, M. M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207(4427), 203–205.
- Kutas, M. M., McCarthy, G., & Donchin, E. E. (1977). Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. *Science*, 197(4305), 792–795.
- Kutas, M., & Federmeier, K. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, 62, 621–647.
- Lang, S., & Kotchoubey, B. (2002). Brain responses to number sequences with and without active task requirement. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, 113(11), 1734–1741.
- Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, 8, 213. <http://dx.doi.org/10.3389/fnhum.2014.00213>.
- Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.
- Luck, S. J., Vogel, E. K., & Shapiro, K. L. (1996). Word meanings can be accessed but not reported during the attentional blink. *Nature*, 383(6601), 616–618.
- Makeig, S., Delorme, A., Westerfield, M., Jung, T.-P., Townsend, J., Courchesne, E., et al. (2004). Electroencephalographic brain dynamics following manually responded visual targets. *PLoS Biology*, 2(6), e176.
- Makeig, S., Westerfield, M., Jung, T., Covington, J., Townsend, J., Sejnowski, T., et al. (1999). Functionally independent components of the late positive event-related potential during visual spatial attention. *Journal of Neuroscience*, 19(7), 2665–2680.
- Marathe, A., Ries, A. J., & McDowell, K. (2013). A novel method for single-trial classification in the face of temporal variability. In D. D. Schmorow & C. M.

- Fidopiastis (Eds.). *Augmented Cognition, HCII 2013. LNCS (LNAI)* (Vol. 8027, pp. 345–352). Heidelberg: Springer.
- McCarthy, G., & Donchin, E. (1983). Chronometric analyses of human information processing. In A. W. K. Gaillard & W. Ritter (Eds.), In G. E. Stelmach & P. A. Voon (Eds.). *Tutorials in event-related potential research: Endogenous components. advances in psychology* (Vol. 10, pp. 251–268). Amsterdam: North Holland Publishing Co.
- Molinero, N., Barber, H., & Carreiras, M. (2011). Grammatical agreement processing in reading: ERP findings and future directions. *Cortex*, 47(8), 908–930.
- Moreno, E. M., & Rivera, I. C. (2013). Setbacks, pleasant surprises and the simply unexpected: Brainwave responses in a language comprehension task. *Social Cognitive and Affective Neuroscience*. <http://dx.doi.org/10.1093/scan/nst066>.
- Münter, T. F. T., Heinze, H. J. H., Matzke, M. M., Wieringa, B. M. B., & Johannes, S. S. (1998). Brain potentials and syntactic violations revisited: No evidence for specificity of the syntactic positive shift. *Neuropsychologia*, 36(3), 217–226.
- Nakano, H., Saron, C., & Swaab, T. Y. (2010). Speech and span: Working memory capacity impacts the use of animacy but not of world knowledge during spoken sentence comprehension. *Journal of Cognitive Neuroscience*, 22(12), 2886–2898.
- Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. (2005). Decision making, the P3, and the locus coeruleus-norepinephrine system. *Psychological Bulletin*, 131(4), 510–532.
- Nieuwland, M. S., & Van Berkum, J. J. A. (2008). The interplay between semantic and referential aspects of anaphoric noun phrase resolution: Evidence from ERPs. *Brain and Language*, 106(2), 119–131.
- O'Connell, R. G., Dockree, P. M., & Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nature Neuroscience*, 15(12), 1729–1735. <http://dx.doi.org/10.1038/nn.3248>.
- Osterhout, L. (1997). On the brain response to syntactic anomalies: Manipulations of word position and word class reveal individual differences. *Brain and Language*, 59(3), 494–522.
- Osterhout, L., & Hagoort, P. (1999). A superficial resemblance does not necessarily mean you are part of the family: Counterarguments to Coulson, King and Kutas (1998) in the P600/SPS-P300 Debate. *Language and Cognitive Processes*, 14(1), 1–14.
- Osterhout, L., & Holcomb, P. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, 31(6), 785–806.
- Osterhout, L., & Nicol, J. (1999). On the distinctiveness, independence, and time course of the brain responses to syntactic and semantic anomalies. *Language and Cognitive Processes*, 14(3), 283–317.
- Osterhout, L., Allen, M. D., McLaughlin, J., & Inoue, K. (2002). Brain potentials elicited by prose-embedded linguistic anomalies. *Memory & Cognition*, 30(8), 1304–1312.
- Osterhout, L., McKinnon, R., Bersick, M., & Corey, V. (1996). On the language specificity of the brain response to syntactic anomalies: Is the syntactic positive shift a member of the P300 family? *Journal of Cognitive Neuroscience*, 8(6), 507–526.
- Perrin, F., García-Larrea, L., Mauguère, F., Bastuji, H., Garcó, L., & Mauguère, É. (1999). A differential brain response to the subject's own name persists during sleep. *Clinical Neurophysiology*, 110(12), 2153–2164.
- Perrin, F., Schnakers, C., Schabus, M., Degueldre, C., Goldman, S., Brédart, S., et al. (2006). Brain response to one's own name in vegetative state, minimally conscious state, and locked-in syndrome. *Archives of Neurology*, 63(4), 562–569.
- Pfefferbaum, A., Ford, J. M., Johnson, R., Wenegrat, B. G., & Kopell, B. S. (1983). Manipulation of P3 latency: Speed vs. accuracy instructions. *Electroencephalography and Clinical Neurophysiology*, 55(2), 188–197.
- Poli, R., Ciel, C., Citi, L., & Sepulveda, F. (2010). Reaction-time binning: A simple method for increasing the resolving power of ERP averages. *Psychophysiology*, 47(3), 467–485.
- Polich, J. (1985). Semantic categorization and event-related potentials. *Brain and Language*, 26(2), 304–321.
- Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, 118(10), 2128–2148. <http://dx.doi.org/10.1016/j.clinph.2007.04.019>.
- Praamstra, P., Meyer, A. S., & Levelt, W. J. (1994). Neurophysiological manifestations of phonological processing: Latency variation of a negative ERP component time-locked to phonological mismatch. *Journal of Cognitive Neuroscience*, 6(3), 204–219.
- Rajkowski, J. (2004). Activation of monkey locus coeruleus neurons varies with difficulty and performance in a target detection task. *Journal of Neurophysiology*, 92(1), 361–371. <http://dx.doi.org/10.1152/jn.00673.2003>.
- Roehm, D., Bornkessel-Schlesewsky, I., Rösler, F., & Schlesewsky, M. (2007). To predict or not to predict: Influences of task and strategy on the processing of semantic relations. *Journal of Cognitive Neuroscience*, 19(8), 1259–1274.
- Roth, W. T., Ford, J. M., & Kopell, B. S. (1978). Long-latency evoked potentials and reaction time. *Psychophysiology*, 15(1), 17–23.
- Rousselet, G. A., & Pernet, C. R. (2012). Improving standards in brain-behavior correlation analyses. *Frontiers in Human Neuroscience*, 6, 119. <http://dx.doi.org/10.3389/fnhum.2012.00119>.
- Roye, A., Jacobsen, T., & Schröger, E. (2007). Personal significance is encoded automatically by the human brain: An event-related potential study with ring tones. *European Journal of Neuroscience*, 26(3), 784–790.
- Royle, P., Drury, J. E., & Steinhauer, K. (2013). ERPs and task effects in the auditory processing of gender agreement and semantics in French. *The Mental Lexicon*.
- Salisbury, D. F., Rutherford, B., Shenton, M. E., & McCarley, R. W. (2001). Button-pressing affects P300 amplitude and scalp topography. *Clinical Neurophysiology*, 112(9), 1676–1684.
- Sara, S. J., & Bouret, S. (2012). Orienting and reorienting: The locus coeruleus mediates cognition through arousal. *Neuron*, 76, 130–141.
- Sanford, A. J., Leuthold, H., Bohan, J., & Sanford, A. J. S. (2011). Anomalies at the borderline of awareness: An ERP study. *Journal of Cognitive Neuroscience*, 23(3), 514–523.
- Small, S. L. (2008). The neuroscience of language. *Brain and Language*, 106, 1–3.
- Smith, J. L., Barry, R. J., & Steiner, G. Z. (2013). CNV resolution does not cause NoGo anteriorisation of the P3: A failure to replicate Simson et al. *International Journal of Psychophysiology*, 89(3), 349–357. <http://dx.doi.org/10.1016/j.ijpsycho.2013.05.002>.
- Spencer, K. M., Dien, J., & Donchin, E. (2001). Spatiotemporal analysis of the late ERP responses to deviant stimuli. *Psychophysiology*, 38(2), 343–358.
- Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science*, 150(700), 1187.
- Szewczyk, J. M., & Schriefers, H. (2011). Is animacy special? ERP correlates of semantic violations and animacy violations in sentence processing. *Brain Research*, 1368, 208–221.
- Townsend, J., Westerfield, M., Leaver, E., Makeig, S., Jung, T.-P., Pierce, K., et al. (2001). Event-related brain response abnormalities in autism: Evidence for impaired cerebello-frontal spatial attention networks. *Cognitive Brain Research*, 11(1), 127–145.
- Ulrich, R. R., & Miller, J. J. (2001). Using the jackknife-based scoring method for measuring LRP onset effects in factorial designs. *Psychophysiology*, 38(5), 816–827.
- van de Meerendonk, N., Chwilla, D. J., & Kolk, H. H. J. (2013). States of indecision in the brain: ERP reflections of syntactic agreement violations versus visual degradation. *Neuropsychologia*, 51(8), 1383–1396.
- van de Meerendonk, N., Indefrey, P., Chwilla, D. J., & Kolk, H. H. J. (2011). Monitoring in language perception: Electrophysiological and hemodynamic responses to spelling violations. *Neuroimage*, 54(3), 2350–2363. <http://dx.doi.org/10.1016/j.neuroimage.2010.10.022>.
- van de Meerendonk, N., Kolk, H. H. J., Chwilla, D. J., & Vissers, C. T. W. M. (2009). Monitoring in language perception. *Language and Linguistics Compass*, 3(5), 1211–1224. <http://dx.doi.org/10.1111/j.1749-818X.2009.00163.x>.
- van de Meerendonk, N., Kolk, H. H. J., Vissers, C. T. W. M., & Chwilla, D. J. (2010). Monitoring in language perception: Mild and strong conflicts elicit different ERP patterns. *Journal of Cognitive Neuroscience*, 22(1), 67–82.
- van de Meerendonk, N., Rueschemeyer, S.-A., & Kolk, H. H. J. (2013). Language comprehension interrupted: Both language errors and word degradation activate Broca's area. *Brain and Language*, 126, 291–301.
- van Herten, M., Kolk, H. H. J., & Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research*, 22(2), 241–255.
- Verleger, R. (1988). Event-related potentials and cognition: A critique of the context updating hypothesis and an alternative interpretation of P3. *Behavioral and Brain Sciences*, 11(3), 343–356.
- Verleger, R. (1997). On the utility of P3 latency as an index of mental chronometry. *Psychophysiology*, 34(2), 131–156.
- Verleger, R. (2010). Popper and P300: Can the view ever be falsified that P3 latency is a specific indicator of stimulus evaluation? *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, 121(8), 1371–1372. <http://dx.doi.org/10.1016/j.clinph.2010.01.038>.
- Verleger, R., Jaśkowski, P., & Wascher, E. (2005). Evidence for an integrative role of P3b in linking reaction to perception. *Journal of Psychophysiology*, 19(3), 165–181. <http://dx.doi.org/10.1027/0269-8803.19.2.xxx>.
- Viola, F., Thorne, J., Edmonds, B., Schneider, T., Eichele, T., & Debener, S. (2009). Semi-automatic identification of independent components representing EEG artifact. *Clinical Neurophysiology*, 120(5), 868–877.
- Vissers, C. T. W. M., Chwilla, D. J., & Kolk, H. H. J. (2006). Monitoring in language perception: The effect of misspellings of words in highly constrained sentences. *Brain Research*, 1106, 150–163.
- Vissers, C. T. W. M., Kolk, H. H. J., van de Meerendonk, N., & Chwilla, D. J. (2008). Monitoring in language perception: Evidence from ERPs in a picture-sentence matching task. *Neuropsychologia*, 46, 967–982.
- Wagenmakers, E. J., Wetzels, R., Borsboom, D., van der Maas, H., & Kievit, R. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, 7(6), 632–638.
- Wassenaar, M., Brown, C., & Hagoort, P. (2004). ERP effects of subject-verb agreement violations in patients with Broca's aphasia. *Journal of Cognitive Neuroscience*, 16(4), 553–576.
- Woody, C. D. (1967). Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals. *Medical and Biological Engineering*, 5(6), 539–554.